Accepted for publication in “Acta Psychologica”

Note: This is an uncorrected version of an author’s manuscript accepted for publication.

The proofs for this manuscript will be copyedited and reviewed before final publication. During this process, small corrections may be made that could affect the content.

Using more different and more familiar targets improves the detection of concealed information

Kristina Suchotzki*[1], Jan De Houwer[2], Bennett Kleinberg[3], and Bruno Verschuere[3]

[1]Department of Psychology, University of Würzburg, Marcusstr. 9-11, 97080 Würzburg, Germany

[2]Department of Experimental-Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium

[3]Department of Clinical Psychology, University of Amsterdam, Nieuwe Achtergracht 128, 1018 WS Amsterdam, The Netherlands

***Corresponding author:**

Kristina Suchotzki

Experimental Clinical Psychology

Department of Psychology

University of Würzburg

Marcusstr. 9-11

97070 Würzburg, Germany

Tel: +49-(0)931-31-82861

Fax: +49-(0)931-31-82733

kristina.suchotzki@uni-wuerzburg.de

Highlights

- The more participants rely on familiarity, the larger effects in the RT CIT may be.

- Using targets that are familiar to participants enhances the validity of the RT CIT.

- Using more different targets also enhances the validity of the RT CIT.

- This supports the role of familiarity in the RT CIT.

Abstract

When embedded among a number of plausible irrelevant options, the presentation of critical (e.g., crime-related or autobiographical) information is associated with a marked increase in response time (RT). This RT effect crucially depends on the inclusion of a target/non-target discrimination task with targets being a dedicated set of items that require a unique response (press YES; for all other items press NO). Targets may be essential because they share a feature - familiarity - with the critical items. Whereas irrelevant items have not been encountered before, critical items are known from the event or the facts of the investigation. Target items are usually learned before the test, and thereby made familiar to the participants. Hence, familiarity-based responding needs to be inhibited on the critical items and may therefore explain the RT increase on the critical items. This leads to the hypothesis that the more participants rely on familiarity, the more pronounced the RT increase on critical items may be. We explored two ways to increase familiarity-based responding: (1) Increasing the number of different target items, and (2) using familiar targets. In two web-based studies (n = 357 and n = 499), both the number of different targets and the use of familiar targets facilitated concealed information detection. The effect of the number of different targets was small yet consistent across both studies, the effect of target familiarity was large in both studies. Our results support the role of familiarity-based responding in the Concealed Information Test and point to ways on how to improve validity of the Concealed Information Test.

Keywords: Response conflict; Familiarity; Concealed Information Test; Guilty Knowledge; Target; Recollection

1. Introduction

Imagine the following scenario. Three robbers enter a burger restaurant and shout ‘All on the floor!’. They grab the money from the cash register and escape on their motorbikes. A few days later, based upon CCTV images, the police identify a possible suspect. The suspect denies involvement in the crime, and the police therefore ask the suspect to take a Concealed Information Test (CIT, also referred to as Guilty Knowledge Test; Lykken, 1959). The CIT consists of a series of multiple-choice questions on the robbery. The suspect may, for instance, be asked where the robbery took place (Was it a gas station? a clothing store? a burger restaurant? a café? a jewelry store?), how many people were involved (Was it one person? two persons? three persons? four persons? five persons?), what the robbers shouted when entering the store (‘Everybody down!’, ‘Give us the money!’, ‘All on the floor!’, ‘Freeze, this is a robbery!’, ‘No crazy shit!’), and how they escaped (the subway? motorbikes? on foot? by car? by minivan?). Denying involvement and any knowledge about the crime, the overt behavioral response of the suspect is not expected to differ between the crime-related and the irrelevant items (i.e., NO). Rather, some indirect reaction to all items is analyzed. This indirect measure may be an autonomic nervous system activity measure such as skin conductance (Lykken, 1959) or the P300 event-related brain potential (Farwell & Donchin, 1991; Rosenfeld et al., 1988). In an adequately designed test that consists of sufficient questions and alternatives and that uses irrelevant items that are well-matched to the crime-related items, it is unlikely for an innocent suspect to consistently show stronger activity to the crime-related than to the irrelevant items. Rather, such pattern of responding reveals recognition of the crime-related details. The validity of the CIT has been well-established (for meta-analyses see Ben-Shakhar & Elaad, 2003; Meijer, klein Selle, Elber, & Ben-Shakhar, 2014).

In recent years, there is an increased interest in the use of response times (RTs) as the dependent variable that is used to assess the recognition of concealed information. In the RT-based CIT, the effect of interest is a typically observed prolonged NO response to critical compared to irrelevant items (i.e., the RT-CIT effect). In addition to the critical and the irrelevant items, the RT-based CIT uses target items. Target items are a dedicated set of items that require a unique response. Building on the example above, the examinee may be instructed to give a unique response (i.e., YES) to such target items as ‘a post office’, ‘six persons’, ‘Nobody moves!’ and ‘sportscar’, and to answer NO to all other items. With the inclusion of such targets, the validity of the RT-based CIT is high. The meta-analysis of Suchotzki, Verschuere, Van Bockstaele, Ben-Shakhar and Crombez (2017), including 34 studies with 1063 participants, found a large effect for the RT-based CIT (Cohen´s d = 1.297; 95% CI [1.060, 1.535]). Whereas that meta-analysis relied solely upon the RT-CIT effect in individuals with concealed information (i.e., sensitivity), Meijer, Verschuere, Gamer, Merckelbach and Ben-Shakhar, (2016) evaluated the diagnostic efficiency in discriminating individuals with versus without knowledge. Their summary of the results of 11 studies with 981 participants showed a high diagnostic efficiency (area under the receiver operating characteristic curve a = .82).

Initially, it was argued that targets in the CIT procedure simply serve to assure attention to the stimuli (Farwell & Donchin, 1991). Whereas evidence suggests that this may be the case for some CIT measures (e.g., the P300 event-related potential; Rosenfeld, Biroschak, & Furedy, 2006), evidence suggests that in the RT-based CIT, targets may serve a much more important role, and actually drive the RT-CIT effect. Matsuda, Nittono, Hirota, Ogawa and Takasawa (2009), for instance, omitted the target items, and simply asked participants to press the same button for all items. With the omission of the target items, the CIT effect disappeared. The same result was found in the fMRI study of Gamer, Klimecki, Bauermann, Stoeter and Vossel (2012): without target items, the well-established, large RT-CIT effect completely disappeared. In sum, research suggests that the use of target items is vital to the validity of the RT-based CIT.

Why might targets be so important? For a knowledgeable suspect, targets and critical items share an important feature that they do not share with the irrelevant items: Familiarity (Verschuere & De Houwer, 2011; see also Seymour & Schumacher, 2009). While critical items are familiar as the result of their link with the critical event under investigation (e.g., a crime), target items are familiar because they are mentioned in the instructions and the examinee is usually required to memorize them. Importantly, responding on the basis of familiarity allows for fast and accurate responding for most of the stimuli: it not only allows for quick YES responses to (familiar) targets but also for quick NO responses to (unfamiliar) irrelevant items, which together with the targets typically account for 5/6th of the trials. For unknowledgeable suspects, familiarity-based responding also allows for rapid rejection of the critical items. It is only for knowledgeable suspects that familiarity-based responding leads to the wrong response for critical items: knowledgeable suspects are familiar with the critical items (which should lead to a YES response when responding is based on familiarity), but want to deny recognition (requiring a NO response). Resolving this response conflict requires time. Direct support of the role of response conflicts is provided by the observation that critical items are associated with increases in RT and activity in brain regions associated with response inhibition only when the examinee is required to answer NO to the critical items, but not when (s)he may answer YES (Suchotzki, Verschuere, Peth, Crombez, & Gamer, 2015). First evidence for the role of familiarity comes from a study by Lukacs, Kleinberg and Verschuere (2017). These authors reported three experiments in which participants tried to hide autobiographical information such as their country of origin. Participants were tested with a RT-based CIT that did or did not include familiarity-related filler trials, that is, trials on which stimuli were semantically related to the concept “familiarity” (e.g., the word ‘familiar’ and the word ‘unfamiliar’) and required the same binary classification as the other stimuli. It was reasoned and found that familiarity-related filler trials would increase the validity of the CIT, presumably because they would promote participants’ greater reliance on familiarity also for the classification of irrelevant and critical items. While suggestive of the role of familiarity, this study did not include a control condition of fillers unrelated to familiarity, opening the possibility that processes other familiarity-based responding contributed to the effect. Moreover, substantial participant loss (up to half of the sample) limits the protocol’s applicability in applied settings.

In the present study, we explore two new ways to increase participants’ reliance on familiarity in the RT-CIT. First, increasing the number of different target items should increase the diagnostic value for participants to base their decisions on familiarity because familiarity is a shared feature of all target items and therefore is the easiest way of identifying different targets. Second, the use of target items that were already familiar to the participant before the test (from now on: familiar targets) as opposed to otherwise unknown items that are only learned during the test and indicated to be targets (from now on: unfamiliar targets) may also increase the probability that responding will be based on a familiarity judgment. In the present study, participants were asked to conceal their country of origin and their birthdate. We orthogonally manipulated the number of different targets (either 2 or 4) and the familiarity with the targets. To achieve the latter, either two irrelevant items were dedicated to be the targets and participants simply learned them before conducting the CIT, or two familiar items – a country and date that participants indicated to be important for them – were chosen to be the targets. We expected successful detection of the concealed autobiographical information, and that detection would be facilitated when using more targets, and when using familiar targets.

2. Experiment 1

2.1. Method

2.1.1. Participants

The study conformed to the principles expressed in the Declaration of Helsinki. Participants were recruited via the data collection website Crowdflower (http://www.crowdflower.com/), provided written informed consent and received 0.50 $ for their participation. Participants were randomly assigned to one of the four between-subject conditions (two familiar targets, four familiar targets, two unfamiliar targets, four unfamiliar targets). In total, we collected data of 444 participants, of which in nine cases, no data was recorded, most likely due to web-browser issues (see Kleinberg & Verschuere, 2015), resulting in complete data for 436 participants. In order to exclude participants who may have taken the test repeatedly, 6 participants were removed because their IP addresses were identical with the IP address of another participant. Of the remaining participants, 73 were excluded because they had 50 % or less remaining trials per item category after the exclusion of error and RT outlier trials.

The final sample consisted of 357 participants, with a mean age of 33.48 years (*SD* = 9.83; *n* = 242 or 68% male participants). The most common native language was English (17%), followed by Serbian (13%), and many other languages (70%). Participants originated mostly from India (13%), Serbia (6%), Bosnia and Herzegovina (6%), and from 59 other countries. Three per cent of the participants indicated that they have obtained at least elementary school, 23% high school, 6% professional training, 20% college, and 49% university education.

Of the final sample, 108 had been assigned to the 2 unfamiliar target condition (M_{Age} = 34.80; *SD* = 10.71; 67% male participants), 60 had been assigned to the 4 unfamiliar target condition (M_{Age} = 32.15; *SD* = 9.63; 67% male participants), 100 had been assigned to the 2 familiar target condition (M_{Age} = 32.88; *SD* = 9.72; 64% male participants), and 89 had been assigned to the 4 familiar target condition (M_{Age} = 33.45; *SD* = 10.02; 74% male participants). The conditions did not differ in the number of men and women, $X^2(3) = 2.41$, $p = .492$, or age, $F(3, 353) = 1.14$, $p = .334$.

2.1.2. Procedure

The study was advertised as a lie detection test in which participants would have to try to hide their true identity. Participants first filled out demographic questions regarding their age, gender, native language and highest education. Then they were asked to choose their true country of origin and their birthdate (day and month, e.g., 13 April) out of a drop-down menu. Those items were used as probes in the CIT. Additionally, everyone was asked to select two more countries and dates (again, day and month) “that were important to them”. For participants in the two familiar targets condition, one of the chosen countries and one of the chosen dates were used as familiar targets in the CIT. For participants in the four familiar targets condition, both chosen countries and dates were used as familiar targets in the CIT. For participants in both unfamiliar target conditions, target items were randomly selected from a list of dates and countries, excluding the chosen countries and dates.

Participants were then instructed that their task was to conceal their true identity and that they had to pretend to be someone else who had one (two target conditions) or two (four target conditions) other identities. This was followed by a target learning procedure in which targets were presented for at least 30 seconds and afterwards had to be repeated by the participants. In case of mistakes, this procedure was repeated until targets were remembered correctly. Participants were then instructed that they would be asked whether the word stimuli presented in the middle of the screen referred to their identity. Therefore, the question “Is this you?” was presented in the upper middle of the screen during the whole task. In order to conceal their true identity and pretend to be someone else, they had to respond YES to their new identity/identities, and NO to all other words, including those related to their true identity. Responses should be given by pressing the “E” key (i.e., Yes, this is me) or the “I” key (i.e., No, this is not me) on their keyboard.

Participants completed a stepwise practice procedure of three distinct practice phases. In the first practice phase, the next stimulus appeared only upon a keypress response, whereas in the second and third practice phase, the next stimulus appeared after 1500ms if no key was pressed. They could only proceed to the next practice phase (and to the actual test phase) when they made less than 50% of errors on target items in each practice phase, and if there was no indication of random, fast button pressing (max. 20% of RTs shorter than 150ms). In the second practice phase, participants had to respond faster than 800 ms on average in order to proceed to the next phase. The third practice phase was identical to the actual test phase with all previous features as well as a “too slow” warning if there was no response within 800ms. The full task can be found and performed on http://www.lieresearch.com/?page_id=719.

In total, each participant completed 240 test trials. Of those, 160 showed irrelevant items (four different dates and four different countries, randomly selected for each participant from a list of dates and countries, each item presented twenty times), 40 showed probes (the actual country of origin and birthday of the participant, each repeated 20 times), and 40 showed targets. Importantly, those targets differed depending on the condition participants were assigned to. In the 2 unfamiliar target condition, one country and one date were randomly selected from a list of dates and countries, and each item was presented twenty times. In the 4 unfamiliar target condition, two countries and two dates were randomly determined from a list of dates and countries, and each item presented ten times. In the 2 familiar target condition, one country and one date were taken from the items that participants had indicated as important to them, and each item was presented twenty times. In the 4 familiar target condition, both countries and dates were taken from the items that participants had indicated as important to them, and each item was presented ten times.

Items were presented in completely randomized order. Reminder labels for the YES and NO responses appeared on the left and right middle part of the screen. Participants were instructed to respond as fast and correct as possible. If participants did not respond within 800ms, the words ‘Too slow’ were presented for 200ms above the presented item. If participants made an error, the word “wrong” was presented below the presented item for 200ms. The inter trial interval was set to vary randomly between 250, 500, and 750ms. The whole procedure (including practice) took approximately 15 minutes.

2.2. Analysis

Data were analyzed with R (https://www.r-project.org/) and raw data as well as the analysis script can be accessed on https://osf.io/6frge/.

The dependent measure was RT in milliseconds. Before data analysis, trials exceeding the response deadline of 1500ms (0.98% of full data set), trials with behavioral errors (e.g., pressing NO for targets; 5.45% of remaining data) and RT outliers (RTs < 150 ms and RTs > 800 ms; 5.20% of remaining data) were excluded. These exclusion criteria were identical to those of previous web-based CITs (Kleinberg & Verschuere, 2015, 2016; Verschuere et al., 2015; Verschuere & Kleinberg, 2016). As in those previous studies, due to the small percentage of error trials, we decided not to report an analysis of the error rate. Results of this analysis are, however, relatively similar to the RT results and can be accessed on https://osf.io/6frge/. Data were analyzed with a 2 x 2 x 2 repeated measures ANOVA, with item type (probe vs. irrelevant items) as within subject factor and number of different targets (two vs. four) and familiarity (unfamiliar vs. familiar) as between subject factors. Note that because our hypotheses focused on the effect of target number and familiarity on the CIT effect (i.e., probe minus irrelevant items), target items were not included in the analyses. Follow-up analyses were performed with paired *t*-tests and Welch's *t*-tests (Delacre, Lakens, & Leys, in press). For ANOVA effects, n_p^2 was calculated as measure of effect size. For group comparisons, the standardized mean difference *d* was calculated as measure of effect size, with .20, .50 and .80 as thresholds for 'small', 'moderate' and 'large' effects (Cohen, 1988). When computing *d* for dependent samples, we corrected *d* for inter-correlations (Dunlap, Cortina, Vaslow, & Burke, 1996; Morris and DeShon, 2002).

2.3. Results

RT means and standard errors of all eight conditions can be found in Figure 1. The 2 x 2 x 2 ANOVA on the RTs revealed significant main effects of item, $F(1,353) = 455.02$, $p < .001$, $n_p^2 = .56$, and of number of different targets, $F(1,353) = 45.76$, $p < .001$, $n_p^2 = .11$. The main effect of familiarity was not significant, $F(1,353) = 1.77$, $p = .184$, $n_p^2 < .01$. The two predicted two-way interactions were significant: There was a significant two-way interaction between item and familiarity, $F(1,353) = 85.13$, $p < .001$, $n_p^2 = .19$, and a significant two-way interaction between item and the number of different targets, $F(1,353) = 5.33$, $p = .022$, $n_p^2 = .01$. The other two-way interaction of familiarity and the number of target items was not significant, $F < 1.2$. As illustrated in Figure 1, the CIT effect was larger in the familiar target condition ($t(188) = 20.24$, $p < .001$, paired Cohen's $d = 1.47$), compared to the unfamiliar target condition ($t(167) = 9.64$, $p < .001$, paired Cohen's $d = 0.74$. The difference between probe and irrelevant items was also larger in the four targets condition ($t(148) = 14.44$, $p < .001$, paired Cohen's $d = 1.18$), compared to the two targets condition ($t(207) = 13.92$, $p < .001$, paired Cohen's $d = 0.96$).

All significant effects were subsumed under the significant three-way interaction, $F(1,353) = 4.12$, $p = .043$, $n_p^2 = .01$. As can be seen in the effect sizes, the influence of familiarity on the probe-irrelevant difference (i.e., the CIT-effect) was larger than the influence of the number of the targets, with the latter being only significant in the familiar targets condition ($t(174.82) = 2.93$, $p = .004$, Cohen's $d = 0.43$), but not in the unfamiliar targets condition ($t(148.24) = 0.23$, $p = .818$, Cohen's $d = 0.04$). In contrast, the influence of familiarity on the probe-irrelevant difference was present both in the two targets ($t(201.85) = 5.72$, $p < .001$, Cohen's $d = 0.80$) as well as the four targets condition ($t(146.32) = 7.78$, $p < .001$, Cohen's $d = 1.24$).

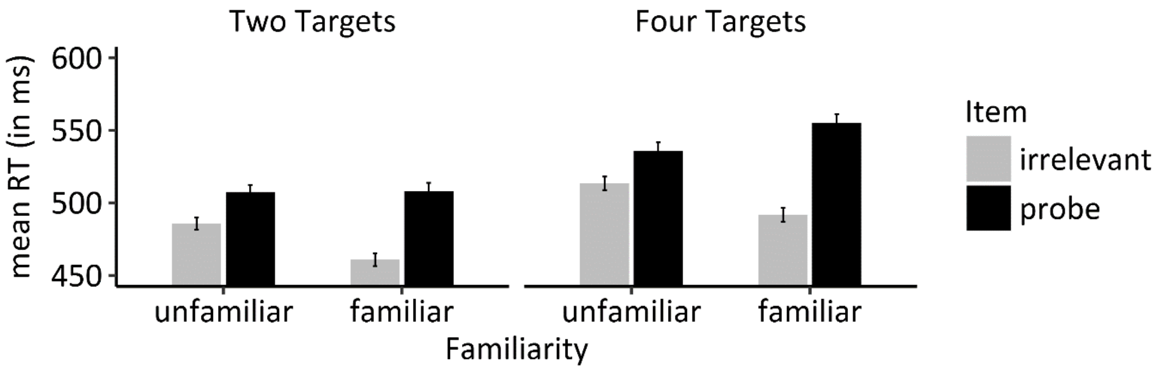

Figure 1. Mean and standard errors of the reaction times in all eight experimental conditions.

2.4. Discussion

Results of the first experiment are in line with our hypotheses. Both increasing the number of target items as well as using target items that were familiar to participants already before the test increased RT differences between probe and irrelevant items. The second experiment was conducted because of two reasons. First, it allowed us to investigate the replicability of both effects (Simons, 2014). Second, due to random variation, the groups in Experiment 1 differed in size. We therefore adapted our randomization procedure with the aim of assigning participants were assigned more equally to the respective experimental conditions.

3. Experiment 2

The main goal of Experiment 2 was to assess replicability of the findings obtained in Experiment 1. To that means, the second experiment was run via the data collection website Prolific (https://www.prolific.ac/). Using *d* of the difference between the probe-irrelevant difference in the two vs. the four target groups to calculate the sample size needed for a power of .95 to replicate this effect, revealed a sample size of 452 participants ($d = .34$, $a = 0.05$, Power = 0.95, calculated with g-power). To compensate for an expected maximum loss of about 37 % (derived from the percentage of data that was lost due to participants not terminating the experiment after the first practice phase and participants that were excluded in the condition with the maximum data loss), we aimed to test 620 participants in total. Evidently, this sample size also provided enough power for the much larger effect of item type x target familiarity ($d = 0.99$; requiring only $n = 56$ participants). The power analysis to compute the sample size, the procedure and hypotheses of our second experiment were preregistered and this preregistration can be found on https://osf.io/q4gwk/register/565fb3678c5e4a66b5582f67.

3.1. Method

3.1.1. Participants

Participants were recruited via the data collection website Prolific (https://www.prolific.ac/), provided written informed consent and received 1.50 £ for their participation. Participants were randomly assigned to one of the four between-subject conditions (two familiar targets, four familiar targets, two unfamiliar targets, four unfamiliar targets). In total, we collected data of 626 participants[1]. In order to exclude participants who may have taken the test repeatedly, 2 participants were removed because their IP addresses were identical with the IP address of another participant. Of the remaining participants, 125 were excluded because they had 50 % or less remaining trials per item category after the exclusion of error and RT outlier trials.

The final sample consisted of 499 participants, with a mean age of 33.03 years (*SD* = 11.50; *n* = 216 or 43% male participants). The most common native language was English (68%), followed by Portuguese (7%), and other languages (25%). Participants originated mostly from the UK (47%) and the US (10%), and from 46 other countries. Three per cent of the participants indicated that they have obtained at least elementary school, 21% high school, 6% professional training, 21% college, and 51% university education. Of those participants, 126 were assigned to the 2 unfamiliar target condition (M_{Age} = 33.65; *SD* = 12.50; 40% male participants), 103 were assigned to the 4 unfamiliar target condition (M_{Age} = 33.81; *SD* = 10.78; 48% male participants), 133 were assigned to the 2 familiar target condition (M_{Age} = 32.34; *SD* = 10.56; 44% male participants), and 137 were assigned to the 4 familiar target condition (M_{Age} = 32.54; *SD* = 11.99; 43% male participants). The conditions did not differ in the number of men and women, $X^2(3) = 1.45$, $p = .695$, or age, $F(3, 495) = 0.52$, $p = .669$.

3.1.2. Procedure and Analysis

The procedure was identical to the one employed in Experiment 1. Data were analyzed with R (https://www.r-project.org/) and raw data as well as the analysis script of Experiment 2 can be accessed on https://osf.io/d2r7s/.[2]

The dependent measure was RT in milliseconds. Before data analysis, trials exceeding the response deadline of 1500ms (1.02% of full data set), trials with behavioral errors (e.g., pressing NO for targets; 5.96% of remaining data) and RT outliers (RTs < 150 ms and RTs > 800 ms; 5.25% of remaining data) were excluded. Due to the small percentage of error trials, an analysis of the error rate is again not reported. Results of this analysis are again relatively similar to the RT results and can be accessed on https://osf.io/d2r7s/. Further data analysis steps were identical to the ones reported for Experiment 1.

3.2. Results

RT means and standard errors of all eight conditions can be found in Figure 2. The 2 x 2 x 2 ANOVA on the RTs revealed significant main effects of item, $F(1,495) = 1056.38$, $p < .001$, $n_p^2 = .68$, and of number of different targets, $F(1,495) = 100.32$, $p < .001$, $n_p^2 = .17$. The main effect of familiarity was not significant, $F(1,495) = 1.08$, $p = .299$, $n_p^2 < .01$. The two predicted two-way interactions were again significant: There was a significant two-way interaction between item and familiarity, $F(1,495) = 160.63$, $p < .001$, $n_p^2 = .25$, and a significant two-way interaction between item and the number of different targets, $F(1,495) = 9.79$, $p = .002$, $n_p^2 = .02$. The other two-way interaction of familiarity and the number of target items was not significant, $F < 2.5$. As illustrated in Figure 2, the CIT effect was larger in the familiar target condition ($t(269) = 29.46$, $p < .001$, paired Cohen's $d = 1.79$), compared to the unfamiliar target condition ($t(228) = 15.59$, $p < .001$, paired Cohen's $d = 1.03$). The difference between probe and irrelevant items was also slightly larger in the four targets condition ($t(239) = 21.35$, $p < .001$, paired Cohen's $d = 1.38$), compared to the two target conditions ($t(258) = 20.00$, $p < .001$, paired Cohen's $d = 1.24$).

All significant effects were subsumed under the significant three-way interaction, $F(1,495) = 13.19$, $p < .001$, $n_p^2 = .03$. As can be seen in the effect sizes, the influence of familiarity on the probe-irrelevant difference was larger than the influence of the number of the targets, with the latter being only significant in the familiar targets condition ($t(267.50) = 4.56$, $p < .001$, Cohen's $d = 0.55$), but not in the unfamiliar targets condition ($t(216.28) = 0.39$, $p = .696$, Cohen's $d = 0.05$). In contrast, the influence of familiarity on the probe-irrelevant difference was present both in the two targets ($t(248.03) = 6.83$, $p < .001$, Cohen's $d = 0.82$) as well as the four targets condition ($t(236.78) = 11.54$, $p < .001$, Cohen's $d = 1.48$).

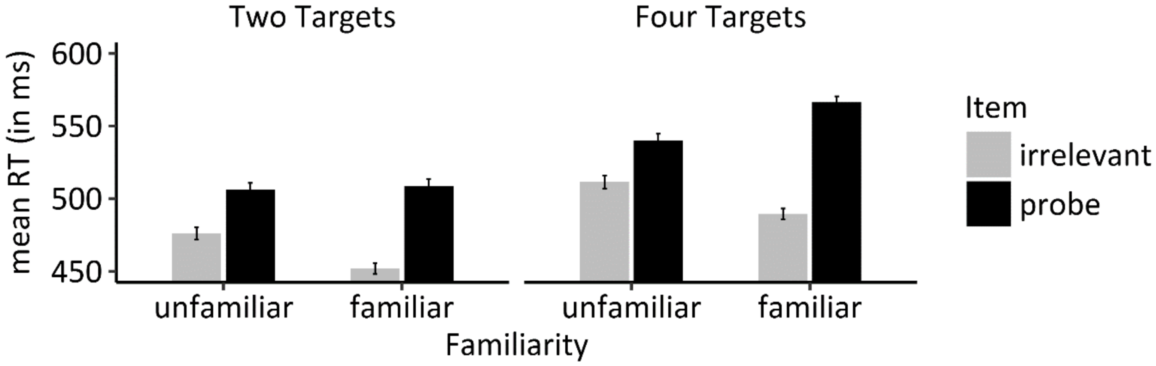

Figure 2. Mean and standard errors of the reaction times in all eight experimental conditions.

3.3. Discussion

The second experiment was run to provide an indication about the replicability of the effects of the number of different targets and of target familiarity on the RT-CIT effects. Before the conduction of the second experiment, we preregistered our procedure and hypotheses and improved our randomization procedure to achieve balanced group sizes. Results of the second experiment very closely replicated the results of the first experiment and are again in line with both hypotheses. Increasing the number of target items as well as using target items that were familiar to participants increased RT differences between probe and irrelevant items. Also, the size of the effects were comparable to the first experiment, with the effect of target familiarity being larger than the effect of the target number.

4. General Discussion

The goal of the present study was to investigate the role of familiarity-based responding in the RT CIT. To that end, we used two manipulations that should increase participants' reliance on familiarity and thereby also increase the CIT effect: Target familiarity and number of targets. Using target items that were familiar to participants already before the test increased RT differences between probe and irrelevant items in comparison to when target items were unfamiliar before the test (effect size Study1: n_p^2 = .19, Study2: n_p^2 =

.25). This is in line with the idea that using familiar targets makes it more advantageous for participants to rely on familiarity, facilitating the fast classification of unfamiliar irrelevant items yet at the same time slowing the correct classification of familiar probe items. A similar yet less pronounced effect was found for our manipulation of the number of targets (effect size Study1: n_p^2 = .01, Study2: n_p^2 = .02). Using more different targets increased RT differences between probe and irrelevant items, possibly also due to fact that it is more advantageous for participants to rely on familiarity when having to identify four different targets compared to having to identify only two different targets. The effect of target number is in line with results of recent RT CIT studies showing that CIT effects seem to be larger and more reliable in test protocols in which multiple probes (plus the corresponding irrelevant items) and thereby also multiple target are presented intermixed, in comparison to protocols in which only a single probe (plus the corresponding irrelevant items) and thereby also only a single target are presented (Lukacs, Kleinberg, & Verschuere, 2017; Verschuere, Kleinberg, & Theocharidou, 2015).

Note that familiarity is not the only possible explanation for the increased CIT effect in the four compared to the two targets condition. A similar prediction could be made based on the hypothesis that imposing cognitive load would increase differences between lying and truth telling (Vrij, Granhag, Mann, & Leal, 2011). Our data also showed a general RT slowing when more targets were used, suggesting a general increase in cognitive load in this condition. The cognitive load explanation does, however, not account for why the effect of the number of targets was especially pronounced in the familiar target condition, as familiar targets should be easier to remember than unfamiliar ones. Moreover, a recent meta-analysis indicates that cognitive load may in fact reduce rather than increase RT differences between lying and truth telling (Verschuere, Köbis, Bereby-Meyer, Rand, & Shalvi, 2017). For the same reason, a general cognitive load account is also unlikely to explain the effect of target familiarity on the RT CIT-effect.

4.1. From theory to practice

Our findings have applied implications. They suggest that by using familiar targets and by increasing the number of targets (in comparison to the number of probes), RT-CIT effects can be further enhanced. The number of targets is most easy to implement in any type of RT-CIT, one just needs to assign a number of irrelevant items to be targets. Our findings suggest, however, that the greatest impact could come from using familiar targets, which may be harder to implement in practice. Referring back to the example of the robbery in the restaurant in the introduction (Was it a gas station? a clothing store? a burger restaurant? a café? a jewelry?), one could for instance use the suspects favorite night club as familiar target location. Or one could use other details of the crime, that have already been disclosed to the suspect during earlier interrogations. Yet such familiar targets may not be easily derived for each question category, for instance in the example mentioned in the introduction, what the robbers shouted when entering the store and how they escaped. Importantly, our data support the conclusion the any measure that increases familiarity-based responding will increase CIT effects, opening up the search for future research for other means to increase familiarity-based responding in CITs.

4.2. Study limitations

The current research also has limitations. First, sample sizes in the four experimental groups were unequal. Although less pronounced in Experiment 2, in both experiments the smallest sample was the one in the four unfamiliar target condition. As participants were randomly assigned to conditions, this is most likely due to a differential drop out rate which in its turn is probably related to the fact that the task is more difficult in the four unfamiliar target condition. This might complicate the applicability of this condition in applied contexts.

Second, an alternative explanation that we cannot exclude with our current design is that instead of familiarity being the crucial feature shared between targets and probes, it is saliency that drives the effects in our familiarity manipulation. Item saliency refers to how important, significant, or relevant something is to the examinee (Dindo & Fowles, 2008) and has been shown to increase the CIT effect (Kleinberg & Verschuere, 2015; Lieblich, Ben-Shakhar, & Kugelmass, 1976). Importantly, while familiarity and saliency will often be strongly related (e.g., one´s own name is highly familiar and highly salient), they could be disentangled experimentally for instance by choosing targets for each participant that they rate before the test regarding both familiarity and saliency and using in the CIT the ones that differ most strongly in those dimensions. Another option would be to experimentally manipulate saliency by making some targets goal-relevant (e.g., by connecting them with a monetary reward) and independently manipulate familiarity by presenting some items more often than others. Third, by using two probes in the CIT, we deviated from the recommended and most effective CIT protocol (with five probes; Ben-Shakhar & Elaad, 2003, see also Verschuere & Kleinberg, 2016). We did this as we wanted to make sure that we could double the amount of target items while still using a number that participants can remember well. Also, in applied contexts it is often not possible to develop a CIT with five probes and CITs with less are common (see e.g., Elaad, 2011). Future research should, however, investigate whether the benefit of increasing the number of target items also hold in a CIT with more probe items.

4.3. Future research

In future research, it would be very interesting to investigate whether effects of target number and familiarity also influence CIT effects in other CIT measures (e.g., skin conductance, heart rate, or neural measures). Recent research suggests that different response measures in the CIT are driven by different processes (klein Selle, Verschuere, Kindt, Meijer, & Ben-Shakhar, 2016; Rosenfeld, Oszan, & Ward, 2017; Suchotzki et al., 2015). Whereas

CIT effects in skin conductance and P300 event-related potentials seem to be mostly driven by attentive orienting towards personally significant information, CIT effects in RTs, heart rate and fMRI measures (e.g., activity in the right inferior frontal gyrus) seem to be driven by the inhibition of the prepotent truthful response. Because one would expect that a stronger reliance on familiarity increases the response conflict for the probe items, one could predict that the number of different target items and their familiarity will mostly affect CIT effects in the latter measures.

Additionally, future research should also include a group of unknowledgeable participants in a crime CIT-design, to investigate whether the use of familiar targets may even also benefit the correct classification of those. Completing the RT CIT with familiar targets may make the test more intuitive and easier for unknowledgeable suspects, as deeper familiarity with the target items could make them easier to distinguish from all irrelevant items (including probes).

4.4. Conclusions

Targets may be essential in the RT-CIT because they share a feature - familiarity - with the critical items. Hence, familiarity-based responding can explain the RT increase on the critical items. This led us to hypothesize that the more participants rely on familiarity, the more pronounced the RT increase on critical items may be. Using familiar targets and using more targets enhances the validity of the RT CIT. These findings not only support the role of familiarity-based responding in the RT CIT but also shows simple yet effective ways to increase CIT effects in applied contexts.

References

Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: A meta-analytic review. *Journal of Applied Psychology, 88*(1), 131-151. doi:10.1037/0021-9010.88.1.131

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale: Lawrence Erlbaum.

Delacre, M., Lakens, D., & Leys, C. (in press). Why psychologists should by default use Welch's t-test instead of Student's t-test. *International Review of Social Psychology*.

Dindo, L., & Fowles, D. C. (2008). The Skin Conductance Orienting Response to Semantic Stimuli: Significance Can Be Independent of Arousal. Psychophysiology, 45, 111-118. http://dx.doi.org/10.1111/j.1469-8986.2007.00604.x

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1*, 170-177. doi:10.1037//1082-989X.1.2.170

Elaad, E. (2011). Validity of the Concealed Information Test in realistic contexts. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 171-186). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511975196.004

Farwell, L. A., & Donchin, E. (1991). The Truth Will Out: Interrogative Polygraphy ("Lie Detection") With Event-Related Brain Potentials. *Psychophysiology, 28*(5), 531–547. doi:10.1111/j.1469-8986.1991.tb01990.x

Gamer, M., Klimecki, O., Bauermann, T., Stoeter, P., Vossel, G. (2012). fMRI-activation patterns in the detection of concealed information rely on memory-related effects. *Social Cognitive and Affective Neuroscience*, *7*, 506-515. doi:10.1093/scan/nsp005

Kleinberg, B., & Verschuere, B. (2015). Memory Detection 2.0: The First Web-Based Memory Detection Test. *PLoS One*, *10*(4). doi:10.1371/journal.pone.0118715

Kleinberg, B., & Verschuere, B. (2016). The role of motivation to avoid detection in reaction time-based concealed information detection. *Journal of Applied Research in Memory and Cognition, 5*(1), 43–51. https://doi.org/10.1016/j.jarmac.2015.11.004

klein Selle, N., Verschuere, B., Kindt, M., Meijer, E. & Ben-Shakhar, G. (2017). Unraveling the roles of orienting and inhibition in the Concealed Information Test. *Psychophysiology*. doi:10.1111/psyp.12583

klein Selle, N., Verschuere, B., Kindt, M., Meijer, E. & Ben-Shakhar, G. (2016). Orienting versus inhibition in the Concealed Information Test: Different cognitive processes drive different physiological measures. *Psychophysiology*, *53*, 579-590.

Lieblich, I., Ben-Shakhar, G., & Kugelmass, S. (1976). Validity of the guilty knowledge technique in a prisoner's sample. Journal of Applied Psychology, 61, 89-93. http://dx.doi.org/10.1037/0021-9010.61.1.89

Lukacs, G., Kleinberg, B., & Verschuere, B. (2017). Familiarity-based fillers improve the validity of reaction time-based memory detection. *Journal of Applied Research in Memory and Cognition*, *6*(3), 295-305. doi:10.1016/j.jarmac.2017.01.013

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology, 43*(6), 385–388. doi:10.1037/h0046060

Matsuda, I., Nittono, H., Hirota, A., Ogawa, T., & Takasawa, N. (2009). Event-related brain potentials during the standard autonomic-based concealed information test. *International Journal of Psychophysiology, 74*(1), 58-68. doi:10.1016/j.ijpsycho.2009.07.004

Meijer, E., Klein Selle, N., Elber, L., & Ben-Shakhar, G. (2014). Memory detection with the Concealed Information Test: A Meta Analysis of Skin Conductance, Respiration, Heart Rate, and P300 data. *Psychophysiology*, 879-904. doi:10.1111/psyp.12239

Meijer, E. H., Verschuere, B., Gamer, M., Merckelbach, H., & Ben-Shakhar, G. (2016). Deception detection with behavioral, autonomic, and neural measures: Conceptual and methodological considerations that warrant modesty. *Psychophysiology, 53*, 593-604. doi:10.1111/psyp.12609

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*(1), 105-125. doi:10.1037//1082-989X.7.1.105

Rosenfeld, J. P., Biroschak, J. R., & Furedy, J. J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *International Journal of Psychophysiology, 60*, 251-259. doi:10.1016/j.ijpsycho.2005.06.002

Rosenfeld, J. P., Cantwell, G., Nasman, V.T., Wojdac, V., Ivanov, S., & Mazzeri, L. (1988). A modified, event-related potential-based guilty knowledge test. *International Journal of Neuroscience, 24*, 157-161.

Rosenfeld, J. P., Ozsan, I., & Ward, A. C. (2017). P300 amplitude at Pz and N200/N300 latency at F3 differ between participants simulating suspect versus witness roles in a mock crime. *Psychophysiology, 54*(4), 640-648. doi: 10.1111/psyp.12823

Seymour, T. L., & Schumacher, E. H. (2009). Electromyographic evidence for response conflict in the exclude recognition task. *Cognitive, Affective, & Behavioral Neuroscience, 9*(1), 71-82. doi:10.3758/CABN.9.1.71

Simons, D. J. (2014). The value of direct replication. Perspectives on Psychological Science, 9(1), 76-80.

Suchotzki, K., Verschuere, B., Peth, J., Crombez, G., & Gamer, M. (2015). Manipulating item proportion and deception reveals crucial dissociation between behavioral, autonomic and neural indices of concealed information. *Human Brain Mapping*, *36*(2), 427-439. doi:10.1002/hbm.22637

Suchotzki, K., Verschuere, B., Van Bockstaele, B., Ben-Shakhar, G. & Crombez, G. (2017). Lying Takes Time: A Meta-Analysis on Reaction Time Measures of Deception. *Psychological Bulletin*, *143*(4), 428-453. doi: 10.1037/bul0000087

Verschuere, B., & Kleinberg, B. (2016). ID-Check: Online Concealed Information Test Reveals True Identity. *Journal of Forensic Sciences*, 62, 237-40. https://doi.org/10.1111/1556-4029.12960

Verschuere, B., & De Houwer, J. (2011). Detecting concealed information in less than a second: response latency-based measures. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 46-63). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511975196.004

Verschuere, B., Kleinberg, B., & Theocharidou, K. (2015). RT-based memory detection: Item saliency effects in the single-probe and the multiple-probe protocol. *Journal of Applied Research in Memory and Cognition*, *4*(1), 59-65. doi:10.1016/j.jarmac.2015.01.001

Verschuere, B., Köbis, N. C., Bereby-Meyer, Y., Rand, D., & Shalvi, S. (2017). Taxing the brain to uncover lying? Meta-analyzing the effect of cognitive load on the reaction time costs of lying. Working paper. Retrieved from https://www.researchgate.net/project/Meta-analysis-on-the-role-of-intuition-vs-deliberation-in-honesty.

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the Liars: Toward a Cognitive Lie Detection Approach. *Current Directions in Psychological Science, 20*(1), 28–32. doi:10.1177/0963721410391245

Footnotes

[1]Note that the randomization procedure was adapted in comparison to the one used in Experiment 1 and also slightly optimized compared to the one described in the preregistration. Instead of creating groups of 31 participants that were alternately assigned to the four experimental groups, smaller groups of 4 or 8 participants (with 1 or 2 per condition, respectively) were created in which the order of conditions was randomized. Due to some technical restrictions of our randomization procedure, there is a slight variation from the pre-registerd sample size per group. We did not collect at least 155 participants per group but 154, 153, 154, and 165 participants.

[2]Note that also the analysis script is identical to the one used in Experiment 1, except for the excluded participants.

Author note

Jan De Houwer is supported by Methusalem Grant BOF16/MET_V/002 of Ghent University.